Big Data and Computing Visions

www.bidacv.com

Big. Data. Comp. Vis. Vol. 2, No. 1 (2022) 18–30.

Paper Type: Original Article

Presenting a Model for the Diagnosis of Heart Failure Using Cumulative and Deep Learning Algorithms: A Case Study of Tehran Heart Center

Amir Hossein Hariri [1], Esmaeil Bagheri[2], Sayyed Mohammad Reza Davoodi[2,*]

[1] Department of Computer Engineering, Technical and Vocational University, Iran; hariri.amirhossein1990@gmail.com.

[2] Department of Computer, Dehaghan Branch, Islamic Azad University, Dehaghan, Iran; bagheri471@gmail.com; smrdavoodi@ut.ac.ir.

Citation:

Hariri, A. H., Bagheri, E., & Davoodi, S. M. R. (2022). Presenting a model for the diagnosis of heart failure using cumulative and deep learning algorithms: a case study of tehran heart center. *Big data and computing visions*, *2*(1), 18-30.

Received: 20/01/2022 Reviewed: 16/02/2022 Revised: 19/02/2022 Accepted: 05/03/2022

Abstract

Coronary artery heart failure is the leading cause of mortality among other cardiac diseases. In most of the cases, angiography is a reliable method for the diagnosis and treatment of cardiovascular diseases. However, it is a costly approach associated with various complications. The significant increase in the prevalence of cardiovascular diseases and the subsequent complications and treatment costs have urged researchers to plan for the better examination, prevention, early detection, and effective treatment of these conditions. The present study aimed to determine the patterns of cardiovascular diseases using integrated classification techniques for analyzing the data of internal medicine patients who are at the risk of heart failure with 451 samples and 13 characteristics. Selecting characteristics and evaluating the influential factors are essential to the development of classifiers and increasing their accuracy. Therefore, we investigated the influential factors of the Gini index. In the classification phase, basic techniques were used, including a decision tree, a neural network, and different cumulative techniques such as gradient boosting, random forest, and the novel deep learning method. A comparison revealed that deep learning with the accuracy of 95.33%, disease class accuracy of 95.77%, and health class accuracy of 94.74% could enhance the presentation and results of the neural network. Out findings confirmed that cumulative methods and selecting influential factors are essential to increasing the accuracy of the diagnostic systems for heart failure. Furthermore, the reported practical tree rules emphasized the use of analytical methods to extract knowledge.

Keywords: Data analysis, Diagnosis of heart failure, Cumulative algorithms, Deep learning.

1 | Introduction

Cardiovascular diseases are non-communicable conditions and a leading cause of mortality worldwide; they have been considered important health conditions since the Middle Ages. Among various cardiac diseases, coronary artery heart failure is particularly associated with a high mortality rate. In most of the cases, angiography is commonly used for the diagnosis and treatment of cardiovascular diseases. However, it is a costly approach associated with numerous complications [1]. Rather than a specific etiology, non-communicable diseases are identified with risk factors, including known social, environmental, and behavioral factors, which increase the risk of a specific disease [2].

Corresponding Author: smrdavoodi@ut.ac.ir

http://dx.doi.org/10.22105/bdcv.2022.325710.1043

The significant increase in the prevalence of cardiovascular diseases and the subsequent costs and complications have urged researchers to plan for the better examination, prevention, detection, early diagnosis, and effective treatment of these conditions. Access to extensive medical data leads to the development of analytical methods for extracting hidden knowledge. Researchers have long been focusing on employing analytical and statistical methods to improve the analysis of large data volumes. Analytical methods have particularly proven efficient for disease diagnosis. Medical data have particular complexities and inconsistencies depending on the type of data. It is not possible to extract a specific pattern from wide-ranging data using conventional methods. Therefore, data mining techniques could interpret medical data more accurately to enhance diagnostic and care processes. Using medical data mining is highly common and plays a key role in health care as it results in the discovery of new, practical, and enduring knowledge in databases [3]. Today, the health sector is in dire need of data analysis, and a testament to this matter is the need to move from traditional medicine to evidence-based medicine.

Diagnosis of cardiovascular diseases has long challenged medical professionals. Therefore, the present study aimed to use data mining techniques to discover and extract the hidden knowledge regarding cardiovascular patients in Al-Zahra Health Center.

The first objective of the research was to investigate factors such as patients' biographic data, such as age, gender, disease course, blood lipids, used medications, stress level, smoking habits, and cardiac disorders. In fact, the main challenge in the research was using algorithms to determine the effects of these factors on cardiovascular diseases by data analysis. Weighting algorithms, such as the Gini index, are such an example, which have been rarely used in previous studies [4].

In the classification stage, the second objective of the study was to use of white-box testing (e.g., a decision tree) to assess and extract practical and hidden rules within data as the output of data mining so that the efficacy of a decision tree and rule generation algorithms. This is because basic techniques (e.g., a decision tree or c4.5) are classic methods that have repeatedly been used in studies and theses [5]. In addition, we will use integrated techniques such as random forest and the new gradient boosting. Recently, using integrated classifications has been on the rise, attracting the attention of machine learning and artificial intelligence researchers. Using a set of weak classifications, these algorithms integrate their outputs to develop the final classification with a higher efficacy than the efficacy of each of the integrated algorithms alone [6].

The first section of this paper provides a review of the literature on cardiovascular disease diagnosis using data analysis. In the second section, the theoretical framework and methodology of the research have been discussed, and the third section focuses on the process of the introduction project. In section four, we have provided the modeling results, including data illustration, pre-processing, and classification. Finally, the paper concludes with defining the evaluation parameters and evaluation results.

2 | Research Background

Cardiovascular disease is one of the leading causes of death and complications in the world. In analyzing clinical data, predicting heart disease is a major challenge. Data mining converts huge amounts of raw data generated by the health industry into useful information that can help make informed decisions.

Machine learning and data mining techniques have recently been used frequently to identify diseases. In [11], an IHDPS was developed using various data mining techniques, including a decision tree, the Naïve Bayes algorithm, and a neural network. According to the findings, each technique had unique potency in meeting the research objectives. The IHDPS could address complex questions of 'What if?", which had not been answered by other decision-making systems.

In [14], a classification approach was proposed for the diagnosis of cardiovascular diseases based on the Naïve Bayes algorithm. In this system, medical data were classified into five categories, including zero risk,

low risk, moderate risk, high risk, and very high risk. Moreover, the system classified the input data into different categories to predict the risk of disease prognosis in a new, unknown sample with 4% accuracy.

Among the learning techniques used to reinforce the basic methods are ensemble learning techniques. These methods take a set of weak classifiers and combine their outputs to make the final classifier in a way that is more efficient than the individual classifier used in the algorithm. The use of methods and techniques such as bagging and boosting and comparing their accuracy, random forest, etiquette enhancement method and many combined techniques are among the important challenges and concepts of data mining among recent articles and dissertations [7].

In [12], a technique was proposed based on cumulative neural networks for the diagnosis of cardiovascular diseases. This model was developed by the integration of the predictive values obtained from previous models. A cumulative approach refers to the integration of multiple neural networks and polling on the prediction results of each neural network model. The accuracy of the model has been estimated at 89.01% based on the assessment of cardiovascular data.

One of the problems of machine learning methods is metozoic data in which due to the small number of samples of disease class (minority class), the accuracy of classification models is challenged.

Alizadeh Sani et al. [15] aimed to diagnose cardiovascular diseases using cost-sensitive algorithms and 10-fold cross-validation on 303 standard data samples. The backup vector machine was reported to have 92.09% accuracy with the optimizer, and the disease classification accuracy (sensitivity) was also estimated at 97.22% [7]. Given the imbalanced disease classification and the healthy population (minimum and maximum), cost-sensitive techniques could be used to assess imbalanced data to increase accuracy.

In [17], a cardiovascular prediction system was employed based on associative classification and a genetic algorithm. The study aimed to classify a high-value technique to integrate a prediction model and reach maximum accuracy in the data obtained from Indian health centers. The applied method proved highly accurate for the diagnosis of cardiovascular diseases. However, feature reduction and selection of influential variables were not involved in the mentioned study.

Arabasadi et al. [1] also proposed a highly accurate integrated technique for the diagnosis of coronary artery diseases based on a reliable dataset. The model could increase efficacy by approximately 10% in a neural network, which used a genetic algorithm to optimize and quantify the neural network weight. Furthermore, feature selection was performed using the Gini index, weighting by the backup vector machine, information benefit, and the analysis of the main components. According to the mentioned study, overall accuracy reached 93.85%, and the comprehensive, integrated technique proposed by the authors was most focused on improving the accuracy of the neural network. As such, the researchers claimed that their method could replace costly angiography.

In [1], the researchers initially used the feature selection method based on correlation and searching particle swarm for a dataset of cardiovascular patients in Indian health centers (335 samples). In addition, k-means clustering was implemented, and the labelled data were classified using a neural network, logistic regression, and the fuzzy rule generation algorithm. Compared to other techniques, model accuracy based on MLP network was reported to be up to 88.4%. The comprehensive technique in the mentioned study applied optimization for feature selection, yielding favorable results. However, the small number of the samples could be considered a limitation, and the researchers considered a larger number of samples for further investigation.

In [8] in 2019, the researchers used 10 data mining classification techniques based on the nearest-neighbor, CART tree, and CHAID tree algorithms, as well as reinforcement methods, for the diagnosis of cardiovascular diseases based on a hospital dataset. The evaluated parameters included patients'

Hariri et al.| Big. Data. Comp. Vis. 2(1) (2022) 18-30

morphological prediction data, blood pressure, perspiration and exercise test, dietary habits, and lifestyle. The results of data mining, which were obtained from decisions trees and AdaBoost, had high accuracy and were recommended for developing disease prediction models. Alizadeh Sani et al. [15] proposed a MVO optimization algorithm using the CAD dataset, which was based on a Multilayer Perceptron (MLP), and the model was compared with nine machine learning methods. Despite 54 features, the researchers used innovative feature selection techniques to enhance the obtained results, and the findings indicated that with the accuracy of 89%, the proposed method could best diagnose coronary artery disease compared to the other nine methods.

Another goal in data mining studies and research is to examine the factors influencing heart attack using feature selection techniques. Various studies have shown that the selection of important features play a key role in improving the performance of machine learning models. In the study [12], Pearson correlation was used to identify the relationship between the characteristics of the heart failure data set. In the prediction step, the nearest neighbor method for classifying healthy and cardiac patients was applied to the data and the results proved that the selection of variables correlated with the target specificity significantly affects the performance of the nearest neighbor (KNN) for the diagnosis of heart failure. Finally, this technique improves the accuracy by up to 97% by selecting the feature.

In 2022, a limited number of authoritative articles have been published on this subject. Study [13] provides a method to investigate the factors affecting heart attacks using the Group Model of Data Handling (GMDH) neural network model. They also compare the results of the proposed method with the results of four neural network models, namely long short-term memory (lstm), probabilistic neural network, redial basis function was compared. The results show that the GMDH technique performs better than the other four methods.

3 | Proposed Model

In the present study, the CRISP-DM method was used. The most successful data mining projects are performed within the framework of standardization, which has been presented by a professional team in the SPSS Inc. as a project named industrial data mining or the Crisp-DM process [13]. The main stages of this process are subject recognition, data interpretation, data clearing and pre-processing, modeling, evaluation, and implementation. *Fig. 3* depicts the process of the current research based on the proposed method and data mining techniques.

3.1 | Data Recognition

The statistical sample of the current research included the dataset of cardiovascular patients, which was collected from various internal medicine centers in Iran. The dataset contained 451 samples and 13 laboratory and clinical patient data, including 261 samples of patients with heart failure and 191 samples of normal patients. According to the information in *Table 1*, the dataset contained one numeric age feature and 12 discrete features. In addition, the dataset was supervised; in other words, it had an objective variable and a classification label. Therefore, the modeling method was selected and analyzed based on this data comprehension. Notably, none of the variables had missing or repetitive values.

3.2 | Data Illustration

At this stage, a histogram of the variables was obtained using statistical data. Normal penetration has been marked with blue, and non-normal penetrations are highlighted with red. Figure 1 illustrates the frequency charts of some features drawn by the researchers. The results of the analysis are as follows:

- *In total, 56% of the patients were male, and 43% were female. Therefore, heart failure more commonly affects men than women.*
- *In total, 30% of the patients had normal weight, 24% were overweight (up to 29 kilograms of excess weight), 26% were obese, and only 17% were underweight.*
- *All the obese patients (over 30 kilograms of excess weight) had heart failure. According to the frequency charts 4-6, increased weight is associated with a higher risk of heart failure. According to the obtained data, 30% of the patients had lower cholesterol levels than 200, while 53% had cholesterol levels of 200-239 (borderline level). The other patients had higher cholesterol levels than 240, and 99% of these patients experienced heart failure.*
- *History of smoking was reported in 52% of the patients, and 80% of these patients had heart failure. In addition, 53% of the patients had a family history of cardiovascular diseases, while no such history was reported in 47%. However, the charts indicated that even the patients without a family history of cardiovascular diseases were at a high risk of developing these disorders, which highlights the role of other risk factors.*
- *Among the patients, 58% had no history of physical exercise, and this category includes the majority of cardiovascular patients. Moreover, 99% of the patients with high diastolic blood pressure had cardiovascular diseases.*
- *Statistical data indicated heart failure even in the patients without a history of hereditary heart failure or smoking habits. Heart failure has also been reported in patients with low cholesterol. Therefore, it could be inferred that statistical data alone may not provide accurate data on the risk factors and diagnosis of cardiovascular diseases, and data mining techniques are required due to overlapping factors.*

Table 1. Dataset variables.

Row	Variable	Type of Variable
	Gender	Discrete Male=1, female=0
	Age	Numeric
1	BMI	Underweight (<18.5)=1 Normal (18.5-24.9)=2 Excess weight (25-29.9)=3 Overweight (>30)=4 Two-value discrete variable
2	Cholesterol	Two-value discrete variable 1=optimal <200 2=between 200-239 3= high above 240
3	Family history	Two-value discrete variable 1= yes 0= no
4	Dyslipidemia	Two-value discrete variable 1= yes 0= no
5	Smoking habits	Two-value discrete variable 1= yes 0= no
6	Exercise	Two-value discrete variable 1= yes 0= no
7	Salt	Two-value discrete variable 1= yes 0= no
8	Stress	Two-value discrete variable 1= yes 0= no
9	Systolic blood pressure	1=hypotension <90 2=desirable=90-119 3=borderline hypertension=120-139 4=hypertension >140
10	Diastolic blood pressure	1= hypotension <60 2=desirable =61-79 3=borderline hypertension =80-89 4= hypertension >=90
11	Heart failure	Two-value discrete variable 1= yes 0= no

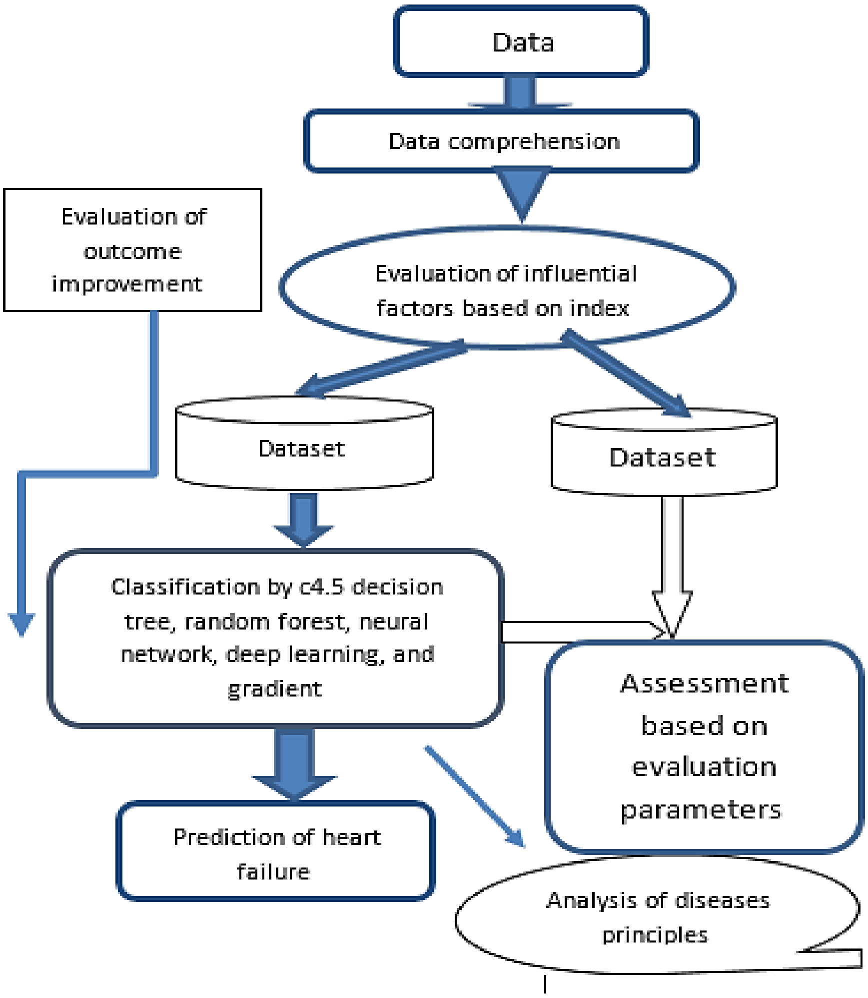


Fig. 1. General research model.

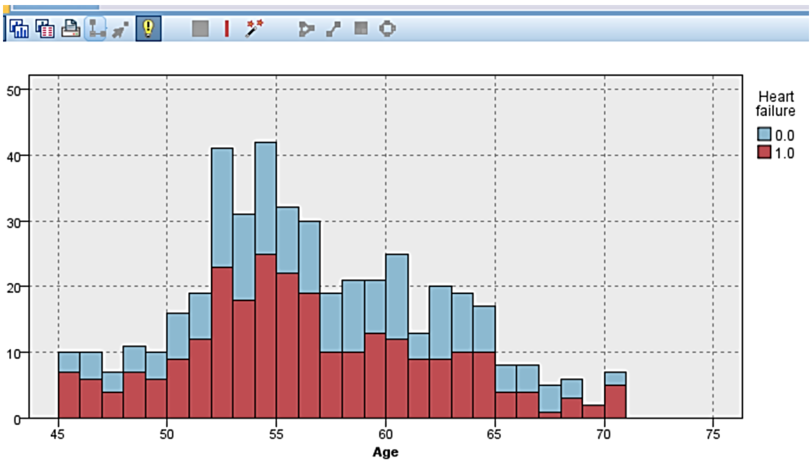
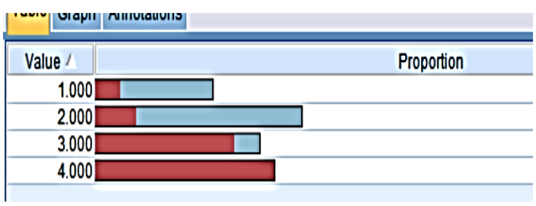
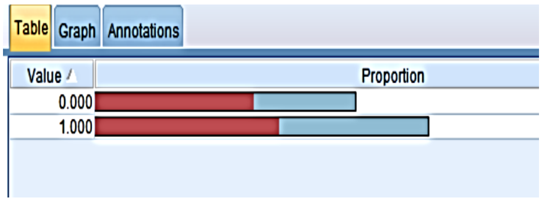


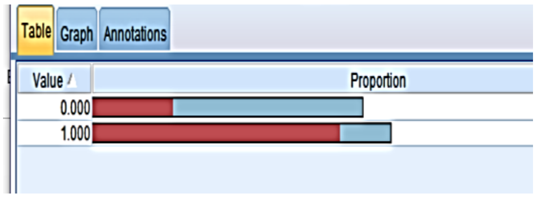
Fig. 1. Frequency of patients' age variable.



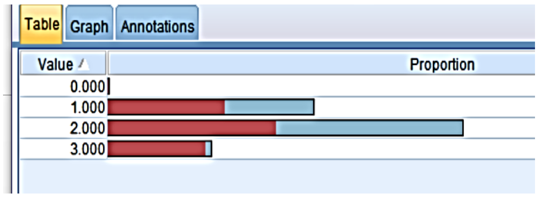
a.



b.

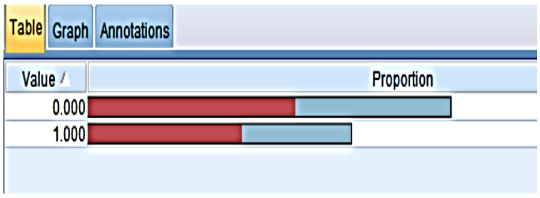


c.

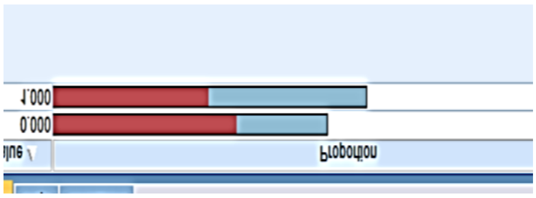


d.

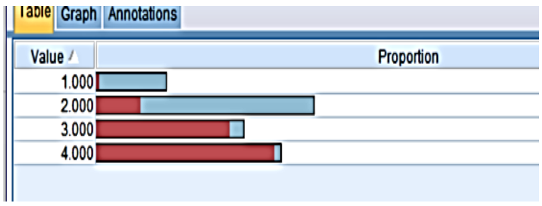
Fig. 2. Index histograms; a. gender, b. BMI, C. cholesterol, D. smoking habits.



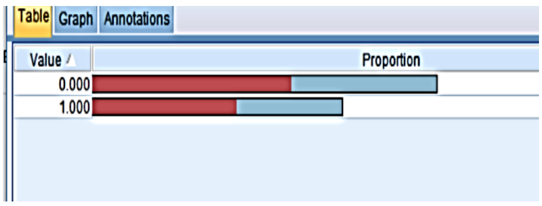
a.



b.



c.



d.

Fig. 3. Index histograms; a. depression, b. inactivity, c. stress, d. diastolic blood pressure.

4 | Modeling (Classification)

In the classification stage of the present study, the results of basic and cumulative classifiers were analyzed. To validate and divided the data into learning and experimental categories, we used smart cross-validation instead of simple validation. After the pre-processing of the dataset, we applied the decision tree, neural network, gradient boosting, random forest, and deep learning classifications to the learning data in the RapidMiner software. After testing, accuracy and the clutter matrix were reported. Each method had specific parameters to adjust the models (*Table 2*). In addition, the optimal values of the parameters were selected meticulously after multiple stages of testing and simulation based on the accuracy of the models.

5 | Evaluation Parameters

Before discussing important classification criteria, it is essential to define a clutter matrix. This matrix shows the performance of a classification algorithm based on the input data, separating various classifications problems. *Table 4* shows the concepts of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN).

Table 2. Parameter adjustment of classifiers.

Abbreviation	Model	Adjustment Parameters
DT	Decision tree	Gini index branches Max. depth: 20 Max. leaf size: 3 Confidence level of pruning: 0.25 Min. records to terminate: 4
MLP	Multilayer perceptron	Recursive neural networks Learning cycles: 500 Momentum coefficient: 0.7 Learning rate: 0.3 Number of layers: 2; first layer: 5, second layer: 6 Active: decay
DL	Deep learning	Number of layers: 50 hierarchical layers Fitness function: hyperbolic tangent

Table 3. Clutter matrix.

Estimated Samples			
Real Samples		Negative Class	Positive Class
	Negative class (-)	TN	FP
	Positive class (+)	FN	TP

- *TP refers to the number of the samples with a positive value (attack) in the test data, which are identified as attacks by the classifier.*
- *FP refers to the number of the samples with a normal value (non-attack) in the test data, which are falsely identified as attacks by the classifier.*
- *TN refers to the number of the samples with a normal value (non-attack) in the test data, which are truly identified as non-attacks by the classifier.*
- *FN refers to the number of the samples with a positive value (attack) in the test data, which are falsely identified as non-attacks by the classifier.*

Evaluation parameters will be explained in the following sections. Classification rate/accuracy is the first criterion in this regard, which shows the percentage of the samples that are classified as positive or negative by the classifier accurately. Based on the clutter matrix, such accuracy is as follows:

$$\text{Accuracy} = \frac{\text{Tp} + \text{TN}}{\text{Tp} + \text{TN} + \text{Fp} + \text{FN}}. \tag{1}$$

Recall or the attack class accuracy (normal class) is another important criterion, in which the numerator is the number of the accurate diagnoses of the positive class (negative), and the denominator is the total number of the samples that are truly positive (negative) in the second line of the clutter matrix.

$$\text{Recall+} = \frac{\text{TP}}{\text{FN+TP}}. \tag{2}$$

$$\text{Recall−} = \frac{\text{TN}}{\text{FP} + \text{TN}}. \tag{3}$$

6 | Discussion and Conclusion

Figs. 4-6 depict the results of the classifications techniques for each implementation stage, as well as the results of overall accuracy and the accuracy of each class.

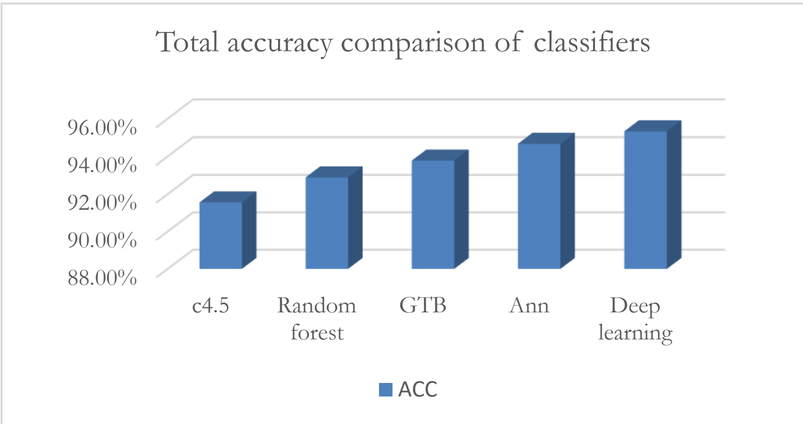


Fig. 4. Total accuracy comparison of classifiers.

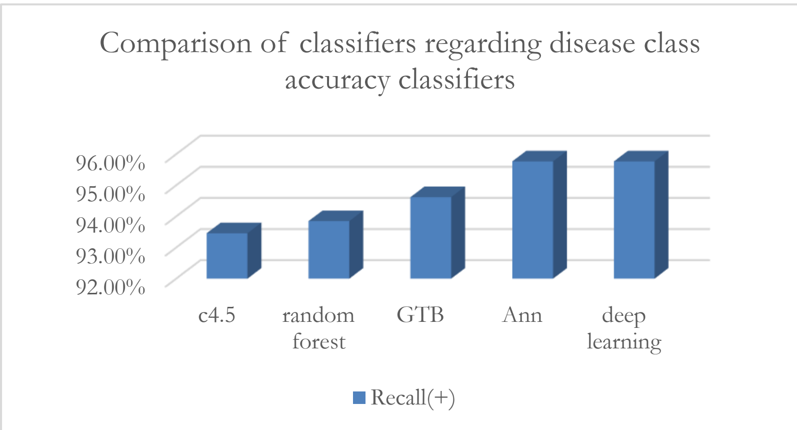


Fig. 5. Comparison of classifiers in terms of disease class accuracy.

Hariri et al.| Big. Data. Comp. Vis. 2(1) (2022) 18-30

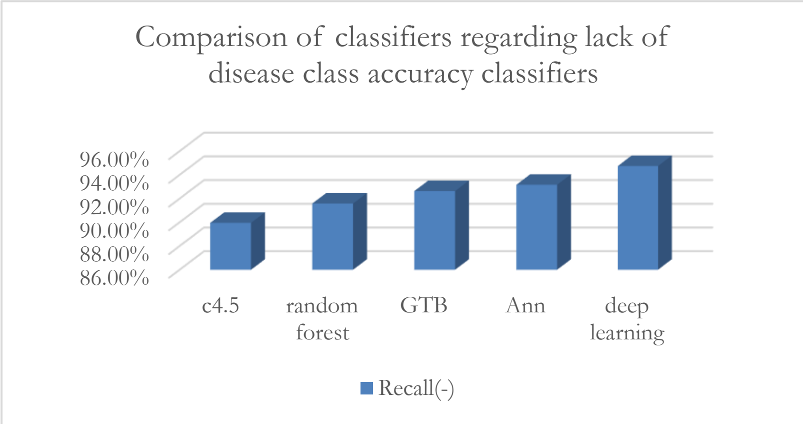


Fig. 6. Normal class accuracy.

- *Deep learning had the highest accuracy (95.33%), while the accuracy of disease class was 95.77%, and the accuracy of health class was 94.74%. In all these cases, excellent outcomes were achieved by the new and improved neural network.*
- *The neural network also yielded favorable outcomes with the accuracy of 94.675%, which was further enhanced by the deep learning algorithm.*
- *On average, the selected algorithms had high accuracy in the diagnosis of hear failure.*
- *Integrated random forest and gradient boosting yielded better outcomes compared to the basic decision tree algorithm. The error rate of this model was improved by eight units in both classes. In addition, gradient boosting yielded significantly better outcomes compared to the random forest algorithm in all three types of accuracy.*
- *In terms of AUC accuracy and the area under the curve of ROC, the horizontal axis reports the negative class error, while the vertical axis shows the accuracy of the positive class. The more the spread chart is inclined toward the vertical axis, the area under the curve is larger, and the model is more successful. The deep learning technique yielded the best outcomes with the area under curve of 0.989, followed by gradient boosting with the accuracy of 0.985.*
- *The significant different in the accuracy of the basic and integrated techniques highlighted the need to use cumulative and integrated methods.*

7 | Analysis of Influential Indexes

To assess the function of each feature and determine their effect on heart failure or lack thereof, we used the Gini index as a weighting data mining index in the software. Conceptually, the Gini index computes variable impurity in proportion to the objective variable. As a result, the weight of each variable is given to the software output through computation. In fact, it is a method used to calculate the impurities of a variable. An unbalanced variable is a variable with the lowest level of impurity, while a balanced knot has the highest level of impurity, being able to distinguish between two classes in classification, thereby providing the most beneficial data [16]. The following equation is used to calculate the profit index using the Gini index:

$$\mathrm{GINI}(t) = 1 - \sum_{j} p(j|t)^2. \tag{4}$$

In the equation above, P(j|t) shows the probability event in proportion to the samples in the j class, which is within the t knot, for the entire samples in this knot. The maximum value in this equation is $(1 - \frac{1}{n})$ since the state of the balanced variable (highest impurity), which is when we have n classes (diseases and lack of diseases classes in our study), partially exists in an equal number of samples from each class regarding the

variable in question, thereby providing most data and profit from the variable. The lowest weight is also observed when more than 80% of the data belong to one class, thereby resulting in low impurity and an imbalanced variable.

We initially report the weight of the indexes. Weights are often within the range of 0-1, and the closer they are to one, they have a more significant effect on the incidence or non-incidence of heart failure. *Table 3* shows the weight of the indexes (variables) based on the entropy index.

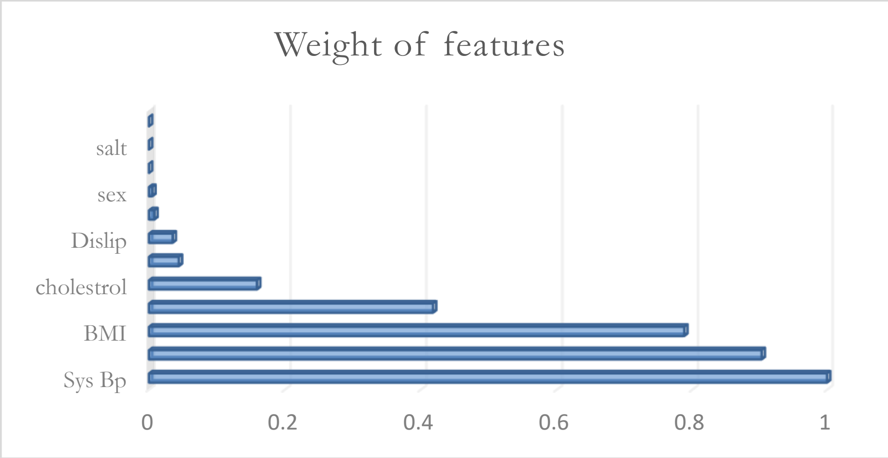


Fig. 7. Weight of variables in dataset.

- *According to the findings, diastolic blood pressure with the weight of one (direct effect) was the most significant influential factor in the diagnosis of heart failure in the health center dataset.*
- *The other influential factors in addition to diastolic blood pressure were the patients' weight, smoking habits, and fat mass.*
- *Variables such as family history and dyslipidemia had moderate effects in this regard.*
- *Stress, high salt intake, and exercise had no significant effects on heart failure (weight: 0).*

8 | Analysis of Rules

Considering the high data accuracy of gradient tree models, we aimed to discuss the production rules and hidden data patterns based on concepts such as penetration detection and normal patterns. *Table 4* indicate that most of the effective rules in this regard are those associated with a higher probability of incidence and a lower error rate.

Table 4. The effective rules for penetration detection.

Number	Rule	Accuracy and Frequency
1	Higher systolic blood pressure than 120 mmHg, normal BMI, and high salt intake increase the risk of heart failure.	100% Observed in 19 patients
2	Normal systolic blood pressure, normal weight, and no history of smoking habits are not associated with a risk of heart failure.	95% Observed in 95 patients
3	Normal systolic blood pressure and normal weight are not associated with the risk of heart failure.	96% Observed in 83 patients
4	Higher systolic blood pressure than 120 mmHg, above-normal weight, and above-normal diastolic blood pressure are associated with the risk of heart failure.	100% Observed in 136 patients
5	Higher systolic blood pressure than 120 mmHg is associated with the risk of heart failure.	93% Observed in 108 patients
6	Blood pressure of 120-139 mmHg and dyslipidemia increase the risk of heart failure.	100% Observed in 36 patients
7	Blood pressure of 120-139 mmHg, absence of dyslipidemia, and high BMI increase the risk of heart failure.	94% Observed in 19 patients
8	Lower blood pressure than 90 mmHg (stage I) is not associated with the risk of heart failure, and the individual is healthy.	89% Observed in 56 patients

Hariri et al.| Big. Data. Comp. Vis. 2(1) (2022) 18-30

9 | Conclusion and Research Implications

In the present study, we aimed to develop a diagnostic system for heart failure using an internal medicine dataset of cardiovascular diseases consisting of 452 data lines, 13 predictive variables, and one objective variable. Based on the findings, basic and cumulative classification techniques were selected. To investigate the effects of the identified influential factors in heart failure, the Gini index was initially implemented on the data, and the effect size of each factor was determined. According to the findings, systolic blood pressure (weight: 1; direct effect), diastolic blood pressure, patient's weight, and history of smoking habits were the most significant influential factors in this regard.

In the classification stage, the data were classified as learning and test data. Two basic techniques (i.e., neural network and decision tree), two cumulative techniques (i.e., gradient boosting and random forest), and the novel deep learning technique were applied to the data. Based on the reported overall accuracy and the results of the clutter matrix, the deep learning method with the accuracy of 95.33%, disease class accuracy of 95.77%, and health class accuracy of 94.74% yielded excellent results.

Significant differences in accuracy between basic and combined methods proved the need to use collective and combined methods. The main purpose of this research is to build a high-precision classification model with deep learning methods and boosting gradient and using the power of these techniques in analyzing heart disease data. Also, the analysis of tree rules and the study of influential factors with the Gini index is another step in extracting knowledge and analyzing the field of heart health using data science, which was discussed in this study.

The significant difference between the accuracy of basic and integrated techniques confirms the need to employ cumulative and integrated methods. For data development, other techniques of influential factor identification are also recommended, such as meta-heuristic approaches. Furthermore, fuzzy decision tree techniques could also be deployed for developing disease rules in future works.

References

[1] Arabasadi, Z., Alizadehsani, R., Roshanzamir, M., Moosaei, H., & Yarifard, A. A. (2017). Computer aided decision making for heart disease detection using hybrid neural network-genetic algorithm. *Computer methods and programs in biomedicine*, *141*, 19-26. https://doi.org/10.1016/j.cmpb.2017.01.004

[2] Maji, S., & Arora, S. (2019). Decision tree algorithms for prediction of heart disease. In *Information and communication technology for competitive strategies* (pp. 447-454). Springer, Singapore. https://doi.org/10.1007/978-981-13-0586-3_45

[3] Jalali, S. M. J., Karimi, M., Khosravi, A., & Nahavandi, S. (2019, October). An efficient neuroevolution approach for heart disease detection. *2019 IEEE international conference on systems, man and cybernetics (SMC)* (pp. 3771-3776). IEEE. DOI: 10.1109/SMC.2019.8913997

[4] Toghraee, M. (2019). Calculation of mean data on gini relationship by data mining method. *Nature, CiiT international journal of data mining and knowledge engineering*, *11*(8), 129-133.

[5] Javeed, A., Zhou, S., Yongjian, L., Qasim, I., Noor, A., & Nour, R. (2019). An intelligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection. *IEEE access*, *7*, 180235-180243. DOI: 10.1109/ACCESS.2019.2952107

[6] Ala'raj, M., & Abbod, M. (2015, September). A systematic credit scoring model based on heterogeneous classifier ensembles. *2015 international symposium on innovations in intelligent systems and applications (INISTA)* (pp. 1-7). IEEE. DOI: 10.1109/INISTA.2015.7276736

[7] Masmoudi, K., Abid, L., & Masmoudi, A. (2019). Credit risk modeling using Bayesian network with a latent variable. *Expert systems with applications*, *127*, 157-166. https://doi.org/10.1016/j.eswa.2019.03.014

[8] Enriko, I. K. A. (2019, June). Comparative study of heart disease diagnosis using top ten data mining classification algorithms. *Proceedings of the 5th international conference on frontiers of educational technologies* (pp. 159-164). https://doi.org/10.1145/3338188.3338220

[9] Rokach, L. (2010). Ensemble-based classifiers. *Artificial intelligence review*, *33*(1), 1-39. https://doi.org/10.1007/s10462-009-9124-7

[10] Wirth, R., & Hipp, J. (2000). CRISP-DM: towards a standard process model for data mining. *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1, pp. 29-39). http://www.cs.unibo.it/~danilo.montesi/CBD/Beatriz/10.1.1.198.5133.pdf

[11] Dangare, Ch. S., & Apte, S. S. (2012). A data mining approach for prediction of heart disease using neural networks. *International journal of computer engineering and technology (IJCET)*, 3(3), 30-40.

[12] Rajasekaran, C., Jayanthi, K. B., Sudha, S., & Kuchelar, R. (2019, July). Automated diagnosis of cardiovascular disease through measurement of Intima media thickness using deep neural networks. *41st annual international conference of the ieee engineering in medicine and biology society (embc)* (pp. 6636-6639). IEEE.

[13] Baselli, G., Cerutti, S., Civardi, S., Lombardi, F., Malliani, A., Merri, M., ... & Rizzo, G. (1987). Heart rate variability signal processing: a quantitative approach as an aid to diagnosis in cardiovascular pathologies. *International journal of bio-medical computing*, *20*(1-2), 51-70.

[14] Joloudari, J. H., Saadatfar, H., GhasemiGol, M., Alizadehsani, R., Sani, Z. A., Hasanzadeh, F., ... & Mansor, Z. (2022). *FCM-DNN: diagnosing coronary artery disease by deep accuracy fuzzy C-means clustering model*. Available at arXiv:2202.04645

[15] Alizadehsani, R., Hosseini, M. J., Sani, Z. A., Ghandeharioun, A., & Boghrati, R. (2012). Diagnosis of coronary artery disease using cost-sensitive algorithms. *IEEE 12th international conference on data mining workshops* (pp. 9-16). IEEE.

[16] Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

[17] Amma, N. B. (2012, February). Cardiovascular disease prediction system using genetic algorithm and neural network. *International conference on computing, communication and applications* (pp. 1-5). IEEE.

Hariri et al.| Big. Data. Comp. Vis. 2(1) (2022) 18-30